

**O.L.Silaeva, V.D.Ilyichev**  
**SPEECH IMITATIONS OF BIRDS**

Severtzov-Institute of Ecology and Evolution of Russian Academy of Sciences  
 Leninsky pr. 33, Moscow 119071, Russia  
 Tel: (7-095) 952-35-84; Fax: (7-095) 954-5534  
 E-mail: silaeva.o@mtu.ru

*Speech imitative signals of talking birds are investigated, using computer program «Electronic Ear» in comparison to prototypal human speech. The frequency of fundamental tone, the frequencies of formant' maxima are determined; the global maxima are revealed, and also the estimation of the width of formants' strips is carried out. The recognition of imitations on hearing, both of the whole word, and of the separate phonemes with the account of co-articulation in a flow of speech and singled-out sounds are estimated also. The comparative legibility of imitations is determined and its connection with the formants' width strips is revealed, in particular, with the strips' width of the first formant.*

The analysis of sounds was carried out using the computer program "Electronic Ear". The program is developed at the Research Center of Image Recognition and is intended for the analysis of speech signals. In the Laboratory of Ecology and Avian Behavior Control, Severtzov-Institute a method have been developed for the analysis of other voice-frequency sounds reproduced by man and animals. The formants analysis of sounds included definition of fundamental tone (FT) frequency, frequencies of formant' maxima (FM), revealing of a global maximum (GM<sup>1</sup>), and also estimation of the width of FM strips. The sound signals were investigated on oscillograms, sonograms, spectrograms and diagrams of FM movements (tracks).

For the analysis of speech signals the program Sonic Foundry Sound Forge 6.0 was used also.

Prototypal human speech signals and speech signals of talking birds were investigated and compared. The comparison of birds' imitations with prototypal speech signals of their owners and trainees whenever possible was carried out.

The follow speech signals of Psittaciformes were investigated:

- Budgerigar, *Melopsittacus undulatus* - 7 individuals
- Cocatiel, *Nymphycus hollandicus* - 1 individual

and Passeriformes:

- Indian Hill Mynah, *Gracula religiosa* - 2 individuals
- Starling, *Sturnus vulgaris* - 1 individual
- Canary, *Serinus canaria* - 1 individual

The records of 6 Budgerigars and one Cockatiel were made by the authors from 1991 to 1996 using the tape recorders "Lira", "Orbita" (microphone MDO 1 "Octave"), "Reporter - 6" (complete microphone) and cassette tape recorders "Electronics - 302" (microphone MDO 1 "Octave"), "Realistic" (microphone Uni-directional dynamic). The parakeets belonged to Moscow birds' amateurs. The record of English talking Indian Hill Mynah was taken from the BBC Sound Archive, the record of Russian talking Indian Hill Mynah, Canary and Starling - from a plate of A. S. Malchevsky, E. N. Golovanova and Y. B. Pukinsky, and Budgerigar Gosha Olevsky - from plates of B. N. Veprinzev. The vowels in a flow of speech were mainly investigated; the attention was given to previous and subsequent consonants also.

The brief analysis results of each signal are given in the tables.

Table 1 did not include the data on the analysis of a Cockatiel' speech signals. It pronounces with singing voice the word "Antonio", it sounds tuneful, but is less legible, than the words, pronounced by B. The investigated sound "o" is tonal and does not contain clicks and explosions unlike the vowels of B. Amplitude modulated signal occupies a wide range of frequencies. FT frequency from 350 Hz at the beginning of the signal increases up to 540 Hz in the end. The general duration of the vowel is about 170 ms.

The first FM occurs in the initial quasiperiods on 1400 Hz (x4), goes upwards on the frequency scale and in last periods occurs on the frequency of 2160 (x4). It is GM. On frequencies of 5000-6000 Hz (x12) there are power splashes, it can be the second poorly expressed FM. The first FM grabs approximately 3 harmonics, that corresponds to the strip width of 1000 Hz.

---

<sup>1</sup> GM – FM with the peak amplitude

**Table 1.** Psittaciformes and their owners

N	Word	announcer	Phone me	FT(Hz)	FM, **GM, width of FM strips (Hz)	Whole duration/ Duration of analyzed part	Number of quasiperiods in analyzed part
1	Anetchka	*B Sok.	«a» first	335	I. 1005(x3) 600-1000 II. <u>1675</u> (x5) 1300-2100	153/57	19
2	Anetchka	Female owner	«a» first	265	I. <u>795</u> (x3) 600-1000 II. <u>1590</u> (x6) 1500-2100	80/56	15
3	Tchay	B Shab.	«a»	180	I. 720(x3) 500-800 II. 1620(x3) 1500-1900 III. <u>2700</u> (x15) 2400-3400	86/86	15
4	Tchay	Female owner	«a»	175	I. <u>700</u> (x4) 500-700 II. 2100(x12) 1900-2100	115/73	13
5	Tchay	B Kit.	«a»	250	I. 1000(x4) 750-1050 II. <u>2000</u> (x9) 1800-2100	82/47	10
6	Tchay	Female owner	«a»	175	I. <u>700</u> (x4) 500-700 II. 1925(x11) 1800-2400	100/95	17
7	Tchayu	B Kal.	«a»	225	I. <u>2475</u> (x11) 2000-3000 II. 7200(x32) 6800-7500	75/75	16
8	Ptitchka	B Sil.	«i»	245	I. <u>2205</u> (x9) 1800-2500 II. <u>3920</u> (x16) 3600-4300	100/79	19
9	Ptitchka	Female owner	«i»	230	I. <u>1460</u> (x2) 400-650 II. 2530(x11) 2000-3000	96/61	14
10	Ptitchka	B Kal.	«i»	215	I. <u>2795</u> (x13) 2500-3000 II. 4085(x19) 3200-4000	137/75	18
11	Ptitchka	Female owner	«i»	165	I. <u>330</u> (x2) 270-370 II. 2640(x16) 2500-3000 III. 4125(x25)	62/62	10
12	Ptitchka	B Olev.	«i»	320	I. <u>2560</u> (x8) 2500-3000 II. <u>3520</u> (x11) 3400-4200	126/70	22

**Table 2.** Passeriformes and their owners

N	Word	announcer	Phone me	FT(Hz)	FM, **GM, width of FM strips (Hz)	Whole duration/ Duration of analyzed part	Number of quasiperiods in analyzed part
1	Charlie	Indian Hill Mynah Charlie	«a»	160	I. <u>800</u> (x4) 650-950 II. 1600(x10) 1500-2500	235/216	34
2	Charlie	Male owner	«a»	95	I. <u>580</u> (x6) 400-600 II. 1140(x12) 1000-1200 III. 1900(x20) 1800-2000	238/232	21
3	Charlie	Female owner	«a»	170	I. 510(x3) 400-600 II. <u>850</u> (x5) 800-1500 III. 2380(x14) 2300-2500	158/128	21
4	Ptitchka	Starling	«i»	250	I. 2250(x9) 2000-2500 II. <u>3000</u> (x12) 2700-3200	372/140	36
5	Sasha	Starling	«a»	420	I. 840(x2) 800-870 II. <u>1260</u> (x3) 1200-1700 III. 2520(x6) 2500-3400	120/64	28
6	Tchika	Indian Hill Mynah Tchika	«i»	490	I. 980(x2) 900-1300 II. <u>2400</u> (x5) 2000-2900 III. 4410(x9) 3900-4800	262/42	21
7	Ptitchki	Canary	«i»	540	I. <u>2700</u> (x5) 2300-2800 II. 5400(x10) 4900-5500	175/34	19

The notes to the tables 1 and 2.

\* B - Budgerigar, near in abbreviated form is given a surname of the owner.

- \*\* GM - is underlined. At alternation GM two FM are underlined.
- A serial number of FM is given in Roman figures
- In brackets after FM value is given repetition factor in relation to FT frequency.

From the table 1 it is visible, that GM at B falls mainly on the second FM, while at human - on the first. Accordingly the greatest energy of a signal is concentrated on frequencies of 2000-2500 Hz. FT frequency considerably varies at different individuals and displays its adjustment to those of the trainees. FM of the vowels at B do not fall outside the limits of 4000 Hz, except the sound [i] from the word "tea" pronounced by B Kal. There is here a distinct FM on the frequency of 7200 Hz. All vowels at investigated B irrespective of a place in a word have power splashes on the frequencies 6000-8000 Hz

The width of FM strips at B (таб.1) varies in large limits in comparison to human ones. However the average value of the strip of the second FM width almost precisely coincides, and is 650 Hz both by B, and by the human, while the strip width of the first FM makes by human about 250 Hz, and by B the same 650 Hz. Probably, it is because of higher values of both FM at B. The tendency of the strips' width of FM increasing is marked at the movement of FM upwards on frequency scale, as occurs at B because of the small size of the voice device. The increasing of the strips width of FM make worse the legibility of speech vowels. An exception is the pronouncing of sound [i] from the word "ptitchka" by B Olev. At the usual for this species wide of FM strips the vowel legibility is very high. B can support FT frequency at the same level as the human beings.

The strip width of the first FM by Passeriformes (таб.2) is more narrow, than by B. The strip width of the first FM changes from 100 up to 500 Hz, on the average it makes 350 Hz against 650 Hz by B. The strip width of the second FM by Passeriformes changes from 500 up to 1000 Hz, on average it makes 750 Hz against 650 Hz by Psittaciformes.

Perception of imitations and prototypal signals. The human hearing completely adequately perceives imitations of birds, at least, investigated here words are easily distinguished. However at perception of sounds isolated from analyzed words, there are some complexities. There are some peculiarities by perception of isolated sounds by different individuals of B. Co-articulation (the influence of a previous and a subsequent consonants on a vowel in the middle) is expressed to a variable degree. This degree depends on a position of the vowel in the word (initial, middle) and from surrounding consonants.

By B Sok. is the legibility very high and the first vowel "a" from the word "Anetchka" sounds clearly and precisely without by-sound "n", but this by-sound is audible in the same sound pronounced by female owner of B Sok. In the word "tchay" and "tchayu" the isolated sound [tʃ] is heard very precisely by all B, it has a high amplitude and large duration in comparison with prototypal sound of the man. The sound "a", isolated from this word, pronounced by B is not to hear at all, an extended crash or click is audible. But the whole word is perceived completely unambiguously. In prototypal human signals "a" sounds precisely, [tʃ] is precise as well, but its amplitude is smaller and the duration is shorter, than in imitative birds' signals.

The sound [i] by birds is heard more precisely, than "a" in the middle of a word. In the word "ptitchka", pronounced by B Sil. the sound [i] is plain, but with clear by-sound [tʃ]. It occurs, probably, because of absence of a pause between these sounds. In the pronunciation of female owner the pause between "i" and [tʃ] makes approximately 20 ms. The by-sound [tʃ] is not present also in the pronunciation of the sound [i] by the female owner of B Kal. in the same word. Here is this sound very short, the pause between "i" and [tʃ] makes approximately 15 ms. It confirms the assumption of the authors, put forward earlier, that the birds, in particular B, succeed better in the articulation of the vowels of the anterior series. However the sound "i" from the same word, pronounced by the Canary, has no by-sound [tʃ]. The reason is probably, that a word is pronounced drawn-out, with a singing voice. The sound [tʃ] has a low amplitude.

The legibility of vowel imitations by Passeriformes is as a whole better, than by Psittaciformes. In isolated vowels the unwanted sounds are not audible. And at the same times the FT frequency by Indian Hill Mynah and Starling can be rather high, but the spectrum does not exceed the bounds of 4800 Hz.

It is possible to assume, that the legibility of vowels is connected with the strip width of FM and especially depends on the strip width of the first FM. Probably, the further the strip width of the first FM is away from 250 Hz (the width of the strip by prototypal human vowels), the less is the legibility of a speech sound.

The authors express their gratitude to the owners of talking birds.